

# SEQ SEARCH SUMMARY

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 10:44:48 ; Search time 2673 Seconds  
(without alignments)  
9015.011 Million cell updates/sec

Title: US-09-602-740-1

Perfect score: 828

Sequence: 1 gaggagcttcgccacatgga.....gcagcgccagctctaacaag 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:  
2: gb\_htg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vi:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_or:  
22: em\_ov:  
23: em\_pat:  
24: em\_ph:  
25: em\_pl:  
26: em\_ro:  
27: em\_sts:

```

28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

%

Result	Query					Description
No.	Score	Match	Length	DB	ID	
1	828	100.0	828	6	AX064875	AX064875 Sequence
2	828	100.0	6995	6	AX074270	AX074270 Sequence
3	828	100.0	6995	6	AX076272	AX076272 Sequence
4	828	100.0	338150	1	AP005278	AP005278 Corynebac
5	828	100.0	349980	6	AX127148	AX127148 Sequence
6	705	85.1	705	6	AX121829	AX121829 Sequence
7	480	58.0	3038	6	AX074275	AX074275 Sequence
c 8	149.6	18.1	15424	1	AE007019	AE007019 Mycobacte
c 9	149.6	18.1	40790	1	MTCY493	Z95844 Mycobacteri
10	134.8	16.3	35881	1	U00013	U00013 Mycobacteri
c 11	134.8	16.3	312050	1	MLEPRTN3	AL583919 Mycobacte
12	104	12.6	1080	6	AX066175	AX066175 Sequence
13	84	10.1	960	6	AX074273	AX074273 Sequence
14	84	10.1	960	6	AX074278	AX074278 Sequence
15	81	9.8	957	6	AX121828	AX121828 Sequence
c 16	73.2	8.8	331	6	AR128446	AR128446 Sequence
17	72.2	8.7	22115	1	SCC22	AL096839 Streptomy
18	61.2	7.4	137478	2	AP004704	AP004704 Oryza sat
19	60.8	7.3	1024	8	AY084434	AY084434 Arabidops
c 20	60.8	7.3	80367	8	AC012329	AC012329 Arabidops
c 21	60.8	7.3	129757	8	ATF2K15	AL132956 Arabidops
22	58.6	7.1	1250	8	AY086161	AY086161 Arabidops
c 23	58.6	7.1	45529	8	AB016884	AB016884 Arabidops
24	57.4	6.9	154029	2	AC120538	AC120538 Oryza sat
c 25	56.8	6.9	148142	1	D90916	D90916 Synechocyst
26	55.4	6.7	10548	1	AE004417	AE004417 Vibrio ch
c 27	55.4	6.7	21149	2	AC013910	AC013910 Drosophil
28	55.4	6.7	99701	2	AC125292	AC125292 Drosophil
29	55.4	6.7	179312	3	AC023745	AC023745 Drosophil
30	55.4	6.7	314087	3	AE003484	AE003484 Drosophil
31	55.4	6.7	334028	2	AC116537	AC116537 Drosophil
c 32	54.6	6.6	10195	1	U32737	U32737 Haemophilus
c 33	54.2	6.5	132685	2	AP005175	AP005175 Oryza sat

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 09:05:53 ; Search time 229 Seconds  
(without alignments)  
8142.597 Million cell updates/sec

Title: US-09-602-740-1

Perfect score: 828

Sequence: 1 gaggagcttcgccacatgga.....gcagcggcagctctaacaag 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query				Description
			Match	Length	DB	ID	
	1	828	100.0	828	22	AAF71360	Corynebacterium gl
	2	828	100.0	6995	22	AAF25332	Nucleotide sequenc
	3	828	100.0	6995	22	AAF29332	C. glutamicum OpcA
	4	828	100.0	349980	22	AAH68529	C glutamicum codin
	5	705	85.1	705	22	AAH66710	C glutamicum codin
	6	480	58.0	3038	22	AAF29319	C. glutamicum OpcA
c	7	149.6	18.1	4411529	22	AAI99682	Mycobacterium tube
	8	104	12.6	1080	22	AAF71023	C. glutamicum SRT
	9	84	10.1	960	22	AAF29320	C. glutamicum OpcA
	10	84	10.1	960	22	AAF29333	C. glutamicum OpcA
	11	81	9.8	957	22	AAH66709	C glutamicum codin
c	12	73.2	8.8	331	21	ABQ62809	Mycobacterium tube
	13	63.4	7.7	53178	23	AAS59543	Propionibacterium
	14	60.8	7.3	1022	21	AAC51469	Arabidopsis thalia
	15	60.8	7.3	1024	21	AAC45134	Arabidopsis thalia
	16	58.6	7.1	1250	21	AAC34656	Arabidopsis thalia
	17	58.6	7.1	1256	21	AAC47798	Arabidopsis thalia
	18	58.6	7.1	1310	21	AAC49386	Arabidopsis thalia
	19	55.4	6.7	732	23	ABL18875	Drosophila melanog
	20	55.4	6.7	2982	23	ABL18874	Drosophila melanog
	21	55.4	6.7	4878	23	ABL04246	Drosophila melanog
	22	54.2	6.5	414	21	AAC98274	Human colon cancer
c	23	53.8	6.5	1010	22	AAI60629	Human polynucleoti
	24	53.8	6.5	1015	22	AAI58843	Human polynucleoti
	25	53.8	6.5	1018	21	AAF21892	Human breast and o
	26	48.2	5.8	6132	21	AAC79584	Virulence gene #4.
	27	47.8	5.8	915	21	AAC41761	Arabidopsis thalia
	28	44.6	5.4	1076	21	AAC49387	Arabidopsis thalia
	29	44.6	5.4	1078	21	AAC35307	Arabidopsis thalia
	30	42	5.1	4966	22	AAS36385	Human cardiovascul
c	31	41.2	5.0	7002	23	ABL09667	Drosophila melanog
c	32	41.2	5.0	10637	23	ABL09666	Drosophila melanog
c	33	41	5.0	6558	21	AAZ45602	cDNA sequence of a
c	34	40.6	4.9	592	21	AAZ93361	Sequence encoding
c	35	40.6	4.9	666	22	AAC84604	Human ZF4 protein
c	36	40.2	4.9	381	23	ABL41662	Nucleotide sequenc
c	37	39.6	4.8	3584	23	ABL07253	Drosophila melanog
c	38	39.6	4.8	13531	23	ABL07252	Drosophila melanog
	39	39.2	4.7	645	17	AAT42476	HuntL4 coding sequ
	40	38.8	4.7	1283	23	AAS85010	DNA encoding novel
	41	38.6	4.7	2853	22	AAI58576	Human polynucleoti
	42	38.6	4.7	2863	22	AAI60362	Human polynucleoti
	43	38.6	4.7	7238	22	AAS36386	Human cardiovascul
	44	38.4	4.6	804	24	ABK48294	Degenerate DNA enc
c	45	38.4	4.6	2224	19	AAV35472	Rat rSK3 coding se

## ALIGNMENTS

RESULT 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 10:54:19 ; Search time 52 Seconds  
(without alignments)  
4883.238 Million cell updates/sec

Title: US-09-602-740-1

Perfect score: 828

Sequence: 1 gaggagctcgccacatgga.....gcagcgccagctctaacaag 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query					Description
No.	Score	Match	Length	DB	ID	
<hr/>						
c 1	149.6	18.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 2	73.2	8.8	331	4	US-09-060-756-308	Sequence 308, App
c 3	40	4.8	2830	4	US-09-221-017B-730	Sequence 730, App
c 4	36.4	4.4	477	4	US-09-135-994-1	Sequence 1, Appli
c 5	36.2	4.4	10348	2	US-08-457-273B-41	Sequence 41, Appli
c 6	36.2	4.4	10348	3	US-08-556-419-13	Sequence 13, Appli
c 7	36.2	4.4	10348	4	US-09-041-886-14	Sequence 14, Appli
c 8	36.2	4.4	10366	1	US-08-246-982A-5	Sequence 5, Appli
c 9	36.2	4.4	10366	1	US-08-453-265-5	Sequence 5, Appli
c 10	35.6	4.3	3819	1	US-07-686-322A-3	Sequence 3, Appli
c 11	35.6	4.3	3819	1	US-08-002-999-3	Sequence 3, Appli
c 12	35	4.2	533	6	5482709-5	Patent No. 5482709

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 11:34:33 ; Search time 58 Seconds  
(without alignments)  
6282.855 Million cell updates/sec

Title: US-09-602-740-1

Perfect score: 828

Sequence: 1 gaggagcttcgccacatgga.....gcagcgccagctctaacaag 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query					Description
	Score	Match	Length	DB	ID	
1	705	85.1	705	9	US-09-738-626-1745	Sequence 1745, Ap
2	81	9.8	957	9	US-09-738-626-1744	Sequence 1744, Ap
3	54.2	6.5	414	10	US-09-925-299-284	Sequence 284, App
4	42	5.1	4966	10	US-09-764-869-1885	Sequence 1885, Ap

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2003, 09:54:28 ; Search time 1967 Seconds  
(without alignments)  
6817.419 Million cell updates/sec

Title: US-09-602-740-1

Perfect score: 828

Sequence: 1 gaggagcttcgccacatgga.....gcagcgccagctctaacaag 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query					Description
	No.	Score	Match	Length	DB	
	1	72.4	8.7	594	14	BQ250641
	2	70.8	8.6	476	10	BE415351
	3	70.8	8.6	490	10	BE443267
	4	70.8	8.6	533	10	BE443748
	5	70.8	8.6	568	10	BE444985
	6	70.8	8.6	578	10	BE443954
	7	70.8	8.6	636	14	BQ805048
	8	70.2	8.5	665	14	BM815572
	9	68.6	8.3	455	12	BG653158
	10	68.6	8.3	601	14	BQ629730
	11	67.8	8.2	807	14	BQ804658
c	12	67.4	8.1	693	10	AW348678
	13	67.4	8.1	1264	10	BE420690
c	14	67	8.1	713	13	BJ263815
c	15	66	8.0	571	10	BE443356
	16	66	8.0	648	13	BJ304570
	17	66	8.0	667	13	BJ280131
c	18	66	8.0	681	13	BJ310981
c	19	66	8.0	701	17	BH535842
	20	66	8.0	787	12	BF259503
	21	66	8.0	796	17	BH694649
	22	65.6	7.9	477	9	AU084921
c	23	65	7.9	654	10	BE444903
	24	64.2	7.8	585	10	AV916596
	25	64.2	7.8	590	10	AV914327
	26	64.2	7.8	633	13	BJ472540
	27	64.2	7.8	654	12	BG415469
	28	64.2	7.8	846	10	BE420795
c	29	63.8	7.7	683	10	AV919353
c	30	63.8	7.7	706	13	BJ474639
	31	63.6	7.7	616	12	BF267119
	32	63.6	7.7	662	14	BQ403858
c	33	63	7.6	571	13	BM074416
c	34	63	7.6	601	13	BM078290
c	35	63	7.6	722	13	BM348739
	36	63	7.6	1316	11	AY105837
	37	62.8	7.6	530	10	AV917228
	38	62.8	7.6	546	10	AV917189
c	39	62.8	7.6	630	10	AV919939
c	40	62.8	7.6	685	10	AV922395
	41	61.6	7.4	805	13	BI311601
	42	61.4	7.4	531	10	AV940134
	43	61.4	7.4	626	13	BI780260
	44	61.2	7.4	581	10	BE194056
c	45	61	7.4	293	14	BQ703295

#### ALIGNMENTS